MEASURING COGNITIVE CHANGE IN ALZHEIMER'S DISEASE CLINICAL DRUG TRIALS

J.E. HARRISON

16 Haygrove Close, Warminster Wiltshire BA12 8SL, United Kingdom, e-mail: john@ceepcee.com

Abstract: In the following paper the cognitive measures used in clinical trials of drugs for Alzheimer's disease are reviewed. The potential benefits of employing innovative tests that map cognitive domains poorly indexed by traditional measures such as the ADAS-cog are considered. Finally, issues pertaining to the cognitive breadth of any proposed new instrument are discussed, as well as the clinical relevance of cognitive change.

Clinical drug trials of putative remedies for Alzheimer's disease (AD) have traditionally sought to capture drug benefits using composite instruments such as the ADAS-cog [1]. The near universal inclusion of this instrument appears to be due to its adoption in early AD clinical drug trials (CDTs), though it is agreed generally that the ADAS is an imperfect instrument and remains so despite efforts to correct its deficiencies. A full critique of the ADAS-cog is beyond the scope of this paper. However, for the purposes of this review it is useful to highlight that the ADAS-cog lacks appropriate measures of key cognitive skills and in particular attention, working memory and executive function. This is by no means a criticism of the original authors of the ADAS-cog. At the time of its introduction the importance of these functions and their susceptibility to decline in even early cases of AD were not well recognized. Some of the original authors have sought to improve the ADAS-cog by adding measures of attention and executive function (e.g. digit cancellation and maze tests, [2]). Experience suggests that many of these recommendations have not been embraced

fully by the drug development community.

If one accepts the premise that the ADAS-cog may not represent the best index of drug efficacy, this begs the question of how best to test the effect of licensed and novel drugs. Irrespective of the indication, measures of cognition should meet the following three prime requirements:
1) The instrument should possess favourable psychometric properties
2) It should be capable of detecting changes in cognition
3) Changes in cognition should be capable of indexing functional status

Selecting tests for neurodegenerative disorders such as AD requires that we also adopt tests sensitive to the stage of the disorder and that we ensure the chosen combination of tests does not exhaust the patient. The necessary psychometric provisions for an objective measure of cognitive function were outlined in Ferris et al.'s [3] valuable paper on this topic. There is very little to add to this important contribution, though later in this paper some recommendations regarding acceptable parameters for test characteristics such as reliability are offered.

Historically the measures employed in CDTs have sought to capture performance in a number of cognitive domains. The first column of Table 1 lists the areas of cognition compromised in patients with AD. Column 2 maps ADAS-cog subtests to these domains and column 3 lists the requirements for cognitive assessment specified by the CPMP in 1997 [4]. Comparison of columns 1, 2 & 3 reveals that, (i) the ADAS-cog does not meet the current guideline requirements and, (ii), that the current guidelines do not specify the assessment of either working memory or executive function. In fairness the guidelines are now a decade old and are currently under revision [5]. Given that both working memory and executive function are recognised as central impairments in AD, it is to be hoped that future guidelines will include a clear requirement to assess these important areas of cognition.

The lack of attentional, working memory and executive function tests in the ADAScog has prompted an increasing number of drug development companies to supplement the ADAS-cog with additional measures. One recent example of this approach was the inclusion of a neuropsychological test battery (NTB) in the Elan/Wyeth AN1792-201 study [6, 7]. Use of this measure, comprised of six well-known cognitive tests, yielded significant benefits of treatment not captured by the ADAS-cog. These encouraging results have prompted other sponsors to adopt versions of the NTB into their own trials.

Recently a number of sponsors have sought to replace the ADAS-cog with assessments based on the NTB. A challenge in achieving this has been the NTB's original focus on memory and executive function. Clearly elements of

Table 1

Cognitive domain	11-item ADAS-cog subtest	CPMP domain	NTB subtest	Supplementary Measures
Episodic verbal memory	Word Recall Word Recognition Remembering test instructions	"Recall and recognition memory…(verbal)"	RAVLT WMS-R VePa	-
Episodic visual memory	-	"Recall and recognition memory…(visuospatial)"	WMS-R ViPa	Rey-Osterreith Complex Figure
Gnosis	Object/finger naming	-	-	Boston Naming Test
Praxis	Constructional praxis Ideational praxis	"constructional ability"	-	Rey-Osterreith Complex Figure
Semantic memory	Orientation	-	-	Orientation*
Language	Language production Word finding difficulty Language comprehension Commands	"language"	COWAT CFT	Commands*
Attention	-	"attention/concentration"	Digit Span	-
Working memory	-	-	COWAT/CFT/DS	-
Executive function	-	-	COWAT/CFT/DS	-
Psychomotor speed	-	"psychomotor speed"	-	Digit cancellation*

RAVLT = Rey Auditory Verbal Learning Test / WMS-R VePa = Wechsler Memory Scale (Revised) Verbal Paired Associates / WMS-R ViPa = Wechsler Memory Scale (Revised) Visual Paired Associates / DS = Digit Span / COWAT = Controlled Oral Word Association Test / CFT = Category Fluency Test; * Indicates that data is borrowed from ADAS-cog subtest

the NTB (see Table 1) necessarily index cognitive skills such as language production and comprehension, planning, reasoning and other faculties. In order to obtain further measures of these functions, several sponsors have augmented the NTB with additional tests, a number of which are listed in column 5 of Table 1.

In selecting measures to either augment or replace the ADAS-cog the drug development community have typically borrowed tests from clinical and experimental psychology. Ultimately it seems probable that we are likely to be best served by cognitive tests developed specifically to characterize the kind of change measurement with which we are most concerned. Many of the available computerized testing systems feature tasks designed for this purpose, though for the most part their use has been limited to early phase clinical development [7, 8]. However, a willingness to use computerized assessments in confirmatory trials is evident from recently published studies of Parkinson's disease dementia [9] and, potentially, future studies of schizophrenia employing the MATRICS battery [10]. Ultimately the utility of the measures we employ must be judged according to their capacity to measure the cognitive constructs compromised in the disease under investigation. Pre-selecting specific tests seems an unnecessary limitation on the measures that could be used, and worse, may seriously limit the opportunities to employ new and potentially better measures.

In terms of objective cognitive test selection, the previously mentioned Ferris paper represents excellent guidance. The authors stressed the importance of good reliability, though did not specify a minimum 'acceptable' level for either internal consistency or test-retest reliability. Kline [11] has suggested that test-retest reliability (TRR) should be >0.7 and internal consistency be in the range of 0.75 to 0.9. In this context it is worth mentioning that the NTB has recorded TRR levels of 0.92 and exhibited desirable levels of internal consistency (Cronbach's alpha = 0.84) [7].

The measures of cognition used in CDTs are typically 'laboratory' measures of cognitive skills and rarely exhibit what Neisser [12] called 'ecological' validity, in that the tasks we employ do not tend to resemble the activities in which we engage in daily life. However, the use of laboratory cognitive tasks should not preclude their use as proxy measures of daily functional activities. In order to interpret the clinical relevance of cognitive preservation or enhancement we need to establish an appropriate benchmark of cognitive change and having established this indicator, to link this to rates of change captured by our chosen measures. Traditionally we have adopted as our benchmark of clinically relevant cognitive preservation/enhancement a 4-point ADAS-cog separation between the treated patients in a CDT when compared to those on placebo. This benchmark has evolved as a consequence of licensing approval issued by

regulators for drugs such as cholinesterase inhibitors. However, there are alternative methods of determining an appropriate benchmark of clinically relevant drug effects. In the next section two approaches are discussed, one based on clinicians' judgements regarding clinical relevance and the other based on the magnitude of statistical effects.

A principled way of determining a benchmark for what has been termed a clinically importance difference (CID) is to ask experienced clinicians. This approach was adopted by Burback et al. [13] who sought to determine whether a consensus CID value could be established with reference to score change on the MMSE [14]. Burback et al.'s study, participated in by 161 Geriatricians and Neurologists, revealed that more than 95% of the responses received specified a CID value of between 2 and 5 inclusive with a mean CID value of 3.72.

Statistics can also provide helpful indicators based on effect size levels deemed to reflect clinical significance. This approach [15] proposes that effect sizes are graded such that less than 0.2 can be regarded as trivial but that greater than 0.2 reflects a 'small' effect, >0.50 a 'medium' effect and >0.8 a 'large' effect. Authors such as Wolf [16] have suggested that effect sizes >0.75 tend to have practical and clinical relevance. Instruments such as the NTB, that correlate well with cognitive change on the MMSE, have the potential to capture clinically relevant preservation or improvement in cognition. It seems likely that the opportunities to establish a clinically relevant effect will be enhanced by the employment of executive function tasks that are known to correlate well with activities of daily living.

The judicious employment of psychometrically robust and drug sensitive cognitive measures has the potential to address issues of both drug safety and efficacy. Preserving the best aspects of traditional tests in combination with the advantages of innovative measures will enable us to enhance the quality of the trials we conduct and allow us to better capture the clinical relevance of the results we obtain.

REFERENCES

1. Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. Am J Psychiat. 141(11):1356-64.
2. Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, Sano M, Bieliauskas L, Geldmacher D, Clark C & Thal LJ. (1997) Development of cognitive instruments for use in clinical trials of anti-dementia drugs: additions to the Alzheimer's Disease Assessment Scale (ADAS) that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer's Dis Assoc Disord. 11:13S-21S.
3. Ferris SH, Lucca U, Mohs R, Dubois B, Wesnes K, Erzigkeit H, Geldmacher D, Bodick N. Objective psychometric tests in clinical trials of dementia drugs. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. Alzheimer Dis Assoc Disord. 1997;11 Suppl 3:34-8.
4. CPMP (1997) Guideline on medicinal products in the treatment of Alzheimer's disease. EMEA: London.
5. CHMP (2005) Recommendation on the need for revision of the guideline on clinical investigation of

medicinal products in the treatment of Alzheimer's disease. EMEA: London.

6. Gilman S, Koller M, Black RS, Jenkins L, Griffith SG, Fox NC, Eisner L, Kirby L, Boada Rovira M, Forette F, Orgogozo J-M. (2005) Clinical effects of A-beta immunization (AN1792) in patients with AD in an interrupted trial. Neurology. 64:1553-1562.

7. Harrison JE, Minassian SL, Jenkins L, Black RS, Koller M & Grundman M (In Press) The NTB: A Neuropsychological Test Battery for use in Alzheimer's disease trials. Arch Neurol.

8. Wesnes KA & Harrison JE (2003). The evaluation of cognitive function in the dementias: methodological and regulatory considerations. Dialogues Clin Neurosci. 5(1): 77-88.

9. Emre M, Aarsland D, Albanese A et al. (2004) Rivastigmine for Dementia Associated with Parkinson's disease. N Eng J Med 351(24): 2509-2518.

10. Marder SR (2006) The NIMH-MATRICS project for developing cognition-enhancing agents for schizophrenia. Dialogues Clin Neurosci. 8: 109-113.

11. Kline P (2000) A Psychometrics Primer. Free Association: London.

12. Neisser, U (1976) Cognition and reality: principles and implications of cognitive psychology WH Freeman: New York.

13. Burback D, Molnar FJ, St John P, Man-Son-Hing M (1999) Key methodological features of randomized controlled trials of Alzheimer's disease therapy. Minimal clinically important difference, sample size and trial duration. Dement Geriatr Cogn Disord. 10(6):534-40.

14. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 12(3):189-98.

15. Cohen J. (1977) Statistical power for behavioural sciences. Academic Press: New York.

16. Wolf FM (1986) Meta-analysis: Quantitative Methods for Research Synthesis. Sage: Beverly Hills.